

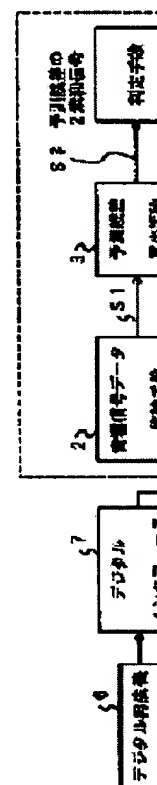
# ATTACK TIME DETECTING DEVICE

Patent number: JP9034448  
 Publication date: 1997-02-07  
 Inventor: SHISHIDO ICHIRO  
 Applicant: VICTOR CO OF JAPAN LTD  
 Classification:  
 - international: G10H1/00; G10G3/04  
 - european:  
 Application number: JP19950205339 19950719  
 Priority number(s):

## Abstract of JP9034448

**PROBLEM TO BE SOLVED:** To provide an attack time detecting device of low cost, a short processing time, and good detecting accuracy.

**SOLUTION:** This attack time detecting device is constituted so that it is provided with an acoustic signal data storing means 2 storing the acoustic signal data of a musical composition, a predictive residual computing means 3 computing predictive residual from the acoustic signal data of the musical composition by a linear prediction method, and a judging means 4 judging the local peak of predictive residual or the local peak of the time variation quantity of predictive residual so as to set the time as the attack time. Hereby, the attack time of good detecting accuracy is obtained in a short time.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-34448

(43) 公開日 平成9年(1997)2月7日

(51) Int.Cl. <sup>8</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 H 1/00			G 1 0 H 1/00	B
G 1 0 G 3/04			G 1 0 G 3/04	

審査請求 未請求 請求項の数 2 F D (全 7 頁)

(21) 出願番号 特願平7-205339

(22) 出願日 平成7年(1995)7月19日

(71) 出願人 000004329

日本ビクター株式会社

神奈川県横浜市神奈川区守屋町3丁目12番地

(72) 発明者 矢戸 一郎

神奈川県横浜市神奈川区守屋町3丁目12番地 日本ビクター株式会社内

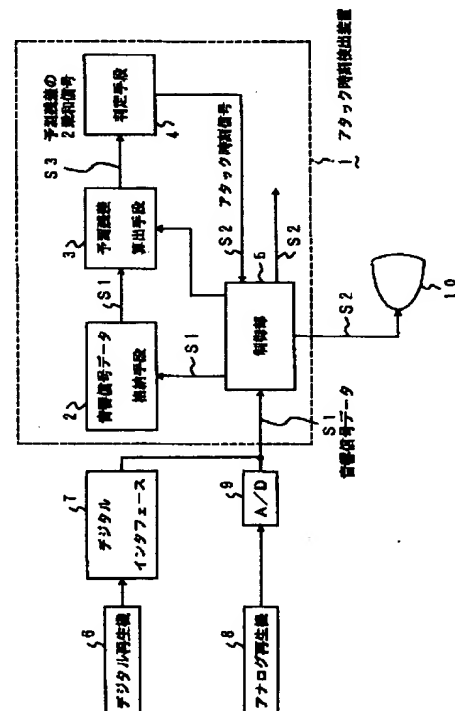
(74) 代理人 弁理士 浅井 章弘

(54) 【発明の名称】 アタック時刻検出装置

(57) 【要約】

【目的】 低コストで処理時間が短く、しかも検出精度の良好なアタック時刻検出装置を提供する。

【構成】 楽曲の音響信号データを格納する音響信号データ格納手段2と、前記楽曲の音響信号データから線形予測法により予測残差を算出する予測残差算出手段3と、前記予測残差の局所ピーク、または前記予測残差の時間変化量の局所ピークを判定してその時刻を楽器音のアタック時刻とする判定手段4とを備えるように構成する。これにより、短時間で検出精度の良好なアタック時刻を求める。



## 【特許請求の範囲】

【請求項1】 楽曲の音響信号データを格納する音響信号データ格納手段と、前記楽曲の音響信号データから線形予測法により予測残差を算出する予測残差算出手段と、前記予測残差の局所ピーク、または前記予測残差の時間変化量の局所ピークを判定してその時刻を楽器音のアタック時刻とする判定手段とを備えるように構成したことを特徴とするアタック時刻検出装置。

【請求項2】 前記予測残差算出手段は、前記音響信号データを一定長のフレーム単位に分割してフレーム内の全てのサンプル或いは一部のサンプルの各予測残差値からそのフレームの予測残差の代表値を算出すると共に、前記判定手段は、フレームの前記予測残差の代表値の局所ピークまたは前記フレームの前記予測残差値の時間変化量の局所ピークを判定して、そのフレームの時刻を楽器音のアタック時刻とするように構成したことを特徴とする請求項1記載のアタック時刻検出装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、楽曲の音響信号から楽譜やMIDIデータを作成する自動採譜装置に関わり、特に各楽器音の立ち上がり時刻（以後、アタック時刻と称す）を検出するアタック時刻検出装置に関する。

## 【0002】

【従来の技術】一般に、シンセサイザや電子ピアノ等の楽器等の音源と連結して情報の交換を可能とするために規格化された情報のフォーマット化方式としてMIDI (Musical Instrument Digital Interface) 規格が、知られている。このMIDI規格により、電子楽器等の駆動信号を効率的に伝達して良好な音質のもとで再生が可能となった。

【0003】ところで、近年通信カラオケ等のように大量のMIDIデータを必要とするシステムが登場しており、これにともなってMIDIデータ制作の効率化が求められている。楽曲の音響信号から自動的にMIDIデータが作成できれば効率の上で理想的であり、これを目的とした自動採譜の研究や特許提案（特開平4-195196号公報など）も行なわれている。しかしながら、現状では、このような技術はごく限られた楽曲に対して

のみ適用可能なレベルなので、実際には人間が聴音または楽譜に頼ってMIDIデータを作成することがほとんどである。

【0004】自動化が困難な理由は、自動採譜の技術において、アタック時刻を検出することが極めて重要であり、且つこの検出が難しいからである。この自動採譜においては、アタック時刻を使って音響信号データを小さな領域に分割し、各領域毎に音程検出等の処理が行なわれる。従って、もしアタック時刻が正確に検出できないと、それ以降の処理に支障を来してしまうことになる。

【0005】楽曲の音響信号データからアタック時刻を検出する装置としては、例えば特開平5-273964号公報が提案されている。これは次に示す手順(a)～(d)を含む技術である。

(a) 楽曲の音響信号データx[n]に対し一定時間間隔（フレーム）毎の対数パワースペクトルの時系列F

[i] [m]（フレーム数I、周波数帯域分割数Mとして、i=1～I、m=1～M）を求める。

(b) F[i] [m]に対する時間変化量の時系列ΔF[i] [m]を求める。

(c) フレーム毎にΔF[i] [m]の正の成分の総和S[i]を数1により算出する。

【0006】

【数1】

$$S[i] = \sum_{m=1}^M \max(\Delta F[i][m], 0)$$

【0007】(d) S[i]の局所ピークを検出し、アタック時刻とする。

【0008】

【発明が解決しようとする課題】ところで、従来装置の例では手順(a)において、数2及び数3を用いて、対数パワースペクトルの時系列F[i] [m]を算出している。これらの数式は従来例に限らず、音響分析において広く使われている計算式であるが、非常に多くの計算量を必要とし、計算時間も長くなる。

【0009】

【数2】

$$O[i][m] = \sum_{k=0}^{2M-1} x[k+120i-M] \cdot \exp(-j \cdot 2\pi(m-1)k/2M)$$

【0010】

【数3】

$$F[i][m] = \log(|O[i][m]|^2)$$

【0011】ただし、i=1～I、m=1～M

この従来装置例では、手順(a)～(d)の処理をCPUを使ったソフトウェア処理で行なっている。ここでは、この装置例で用いられているパラメータ値(M=256)を使って上記数2、3を比較的良く使われる32

ビットCPU(10～20MIPS程度の能力)で計算すると、1フレーム当りおよそ数10msec時間がかかってしまう。この装置例では、音響信号の10msecに1つの割合でフレームを1つ設定しているの、手順(a)だけの処理に実時間のおよそ数倍の時間が必要となり、大量のデータを処理する場合にかなりの時間が必要である。

【0012】勿論、高速なCPUや特殊なハードウェア

を使えば処理時間も少なく済むが、処理装置のコストも高くなる。従って、比較的低コストで、処理時間が短く、しかも検出精度の良いアタック時刻検出装置が望まれていた。本発明は、以上のような問題点に着目し、これを有効に解決すべく創案されたものであり、本発明の目的は、低コストで処理時間が短く、しかも検出精度の良好なアタック時刻検出装置を提供することにある。

#### 【0013】

【課題を解決するための手段】本発明は、上記問題点を解決するために、楽曲の音響信号データを格納する音響信号データ格納手段と、前記楽曲の音響信号データから線形予測法により予測残差を算出する予測残差算出手段と、前記予測残差の局所ピーク、または前記予測残差の時間変化量の局所ピークを判定してその時刻を楽器音のアタック時刻とする判定手段とを備えるように構成したものである。

【0014】このように構成することにより、音響信号データ格納手段に記憶された音響信号データに基づいて、予測残差算出手段は線形予測により、予測残差を算出する。判定手段は、この予測残差に基づいて、この局所ピークや予測残差の時間変化量の局所ピークを判定してそれをアタック時刻とする。これにより、処理時間も少なく、検出精度も高いアタック時刻を求めることが可能となる。

#### 【0015】

【発明の実施の形態】以下に、本発明に係るアタック時刻検出装置の一実施例を添付図面に基づいて詳述する。図1は自動採譜装置等に用いられる本発明に係るアタック時刻検出装置を示すブロック構成図、図2は図1の装置で作成されるフレームを説明する説明図、図3は予測残差算出手段の動作を示すフロー、図4は判定手段の動作を示すフロー、図5はアタック時刻の判定を説明するための説明図である。

【0016】図1に示すようにこのアタック時刻検出装置1は、全体が例えばマイクロコンピュータ等により構成されており、楽曲の音響信号データS1を格納する音響信号データ格納手段2と、線形予測法により上記格納された音響信号データから予測残差を算出する予測残差算出手段3と、この予測残差の局所ピークや予測残差の時間変化量の局所ピークを求めて、その時刻をアタック時刻とする判定手段4とにより主に構成されている。上記各手段の動作は、制御部5によりコントロールされる。上記音響信号データS1は、再生機がCD、DAT (Digital Audio Tape) 等のデジタル再生機6の場合にはデジタルインタフェース7を介して入力され、再生機がアナログテープのようなアナログ再生機8の場合には、所定の周波数、例えば12KHz程度のサンプリング周波数でサンプリングを行なってデジタル信号に変換するA/D変換器9を介して入力される。

【0017】上記判定手段4にて得られたアタック時刻信号S2は、自動採譜装置の本体(図示せず)にてMIDI信号化のために用いられると同時に、ディスプレイ10等にも表示されて、オペレータに知らせようになっている。音響信号データS1は、上述のようにデジタルインタフェース7を介してCD、DATを再生するデジタル再生機6から直接デジタル入力されるか、或いはアナログ信号の場合は、再生されたアナログ信号をA/D変換器8を介してデジタル化し、音響信号格納手段2にデジタルデータとして格納される。音響信号データ格納手段2は、半導体メモリ、磁気ディスク等で構成することができる。

【0018】次に、予測残差算出手段3について説明する。予測残差算出手段3は、音響信号格納手段2より供給される音響信号データを複数のフレームに分割し、各フレーム毎に予測残差の代表値を算出して出力するものである。尚、本実施例では、予測残差の代表値として各データサンプルの予測残差の2乗和を算出する場合について説明する。この予測残差算出手段3の動作フローを図3に示す。まず、格納されている音響信号データを図2に示すようなフレームFに分割してFN個のフレームを作成する(S11)。ここで1フレームは、FL個のデータサンプルで構成され、前後のフレーム間の間隔は、FP個のデータサンプルに相当する。フレーム同士は、一部重なり合っても良いし、また、離れていてもよい。フレームの中央のデータサンプルの時刻をそのフレームの時刻とする。FL、FPの最適値は、サンプリング周波数や楽曲の特性により変わるが、例えばサンプリング周波数12KHzの時にFP=120とすると、10msecという実用上十分な精度でアタック時刻を求めることができる。

【0019】フレームの総数をFNとすると、 $i=1 \sim FN$ のフレームについて以下の処理を行い予測残差の2乗和 $E[i]$ を算出する。まず、 $i$ を1として(S12)、数4に示すように、音響信号データ $x[n]$  ( $n=1 \sim FL$ )に窓関数 $w[n]$ を乗じる(S13)。この窓関数としては数5に示すハミング窓を用いる。

#### 【0020】

【数4】

$$X[n] = x[n] \cdot w[n]$$

【0021】ただし、 $n=1 \sim FL$

【0022】

【数5】

$$w[n] = 0.54 - 0.46 \cos(2\pi n / FL)$$

【0023】窓がけしたデータ $X[n]$ に対して線形予測法に基づく処理を行なう。線形予測法は、信号波形の分析/合成に広く使用されている手法である。ここでは数6に示すように、現在のデータ $X[n]$ をそれ以前の $p$ 個のデータ $X[n-1]$ ,  $X[n-2]$ , ...,  $X[n-p]$ の線形結合と残差 $e[n]$ との和で表し、フレー

ム内の残差  $e[n]$  の2乗和が最小となるような係数  $a[k]$  を算出する。

【0024】

【数6】

$$X[n] = - \sum_{k=1}^p a[k] X[n-k] + e[n]$$

$$E[i] = \sum_{n=1}^{FL} (X[n] + \sum_{k=1}^p a[k] X[n-k])^2$$

【0027】

$$\sum_{k=1}^p a[k] \sum_{n=1}^{FL} X[n-k] X[n-q] = - \sum_{n=1}^{FL} X[n] X[n-q]$$

【0028】ただし、 $1 \leq q \leq p$  以上のS3とS4の動作を、 $i$ の値がFNになるまで(S15)、 $i$ を1ずつインクリメントしつつ繰り返し行い(S16)、残差の2乗和を求める。一般的な線形予測法の使用例では、信号波形をなるべく忠実に表現することを目的とするので、 $p$ の値をある程度大きくして波形情報を $a[k]$ で表現するようにし、 $e[n]$ を極力小さくするようにする。

【0029】一方、本発明では、信号波形を忠実に表現することは目的でなく、フレーム内の予測残差の2乗和を使用するという点で、一般的な使用方法とは異なる。

【0031】

$$a[1] = R[1] (R[2] - R[0]) / (R[0]^2 - R[1]^2)$$

$$a[2] = (R[1]^2 - R[0] R[2]) / (R[0]^2 - R[1]^2)$$

$$\text{ただし、} R[q] = \left( \sum_{n=1+q}^{FL} X[n] X[n-q] \right) / (N-q), q=0, 1, 2$$

【0032】また、この場合の予測残差の2乗和  $E[1]$  は、数7、数8を変形して自己相関関数を使って表した数11により算出する。

$$E[i] = R[0] + a[1] R[1] + a[2] R[2]$$

【0034】 $p$ を他の値にする場合も全く同様に数7、数8を変形して予測残差の2乗和  $E[i]$  を求めることができる。以上の処理により、フレームの予測残差の2乗和が求まるが、この値が大きい場合は、フレーム内で音響信号データの変動が激しく、この値が小さい場合は、変動が少ないことになる。アタックのある場所では、音響信号データは急激に変化するので、フレーム内の予測残差の2乗和も大きくなる。従って、そのようなフレームの時刻をアタック時刻とすれば良い。

【0035】また、本発明では、フレーム内の複数のデータサンプルから得られる複数の予測残差値から、そのフレームの予測残差の代表値を算出して使用することにより、アタック時刻の検出誤差を小さくしている。従って、この予測残差の代表値は、予測残差の2乗和に限ら

【0025】 $i$ 番目のフレームの残差の2乗和  $E[i]$

は、数7で表わすことができ、これを求める(S14)。これを最小とする条件として、係数  $a[k]$  に関する偏微分を0とおくと、数8が得られる。

【0026】

【数7】

【数8】

データサンプル毎の  $e[n]$  そのものではなく、フレーム内の予測残差の2乗和を使うので、 $p$ の値をかなり小さくすることが可能となる。 $p$ を小さくすることにより、計算時間が短縮できるのは言うまでもない。例えば1フレーム中のデータサンプル数  $FL$  が数100程度の場合でも、 $p=2$ 程度のアタック時刻の検出が可能である。 $p=2$ の場合の係数  $a[1]$ 、 $a[2]$ の値は、数8を変形し、自己相関関数  $R[0]$ 、 $R[1]$ 、 $R[2]$ を使って表した数9、数10により算出する。

【0030】

【数9】

【数10】

【0033】

【数11】

ず、予測残差の2乗和の平方根や絶対値の総和などを予測残差の代表値として後の処理で使用しても良い。また、フレーム内の全てのサンプルの予測残差の2乗和を求めるのではなく、例えばフレームの中央部のいくつかのサンプルに対して予測残差の2乗和を求め、これを後の処理で使用しても良い。

【0036】予測誤差算出手段3は、従来装置例の手順(a)に相当するが、従来装置例に比べると計算量が少なく済む。従って、予測誤差算出手段3を汎用CPUを使ったソフトウェア処理で実現した場合に、従来装置例に比べて処理時間が短くて済む。例えばフレームデータサンプル数  $FL=240$ 、フレーム間のデータサンプル相当数  $EP=120$ 、サンプリング周波数  $12\text{ KHz}$  として、従来装置例と同じくフレームの間隔を  $10\text{ ms}$

ecと仮定すると、前述した性能のCPUで計算した場合には、1フレーム当り約数msecとなり、従来方式の約1/10の時間で計算することができる。

【0037】判定手段4においては、図4に示すフローに基づいて処理を行なう。まず、予測残差の2乗和E[i]の時間変化量 $\Delta E[i]$ を求める。 $\Delta E[i]$ の算出式としては種々のものが利用できるが、例えば数12を用いることができる。

【0038】

【数12】

$$\Delta E[i] = E[i+1] - E[i-1]$$

【0039】ただし、 $i=1 \sim FL$ 、 $E[-1] = 0$ 、 $E[FL+1] = 0$

$$\Delta E[i] > \Delta E[i-1] \cap \Delta E[i] > \Delta E[i+1] \cap \Delta E[i] > \alpha$$

【0042】ただし、 $\alpha$ はある定数であり、閾値を示す。まず、上記数13において $i=1$ として(S25)、局所ピークの判定を行なう(S26)。数13から明らかなように、 $i$ を1つずつインクリメントしながら連続する3つの時間変化量 $\Delta E[i]$ の大小関係を比較する(S27、S28)。そして、 $i$ の値がFNになったならば処理を終える(S27のNO)。

【0043】これにより、図5に示すようにアタック時刻を求めることができる。尚、楽曲によっては、 $\Delta E[i]$ の算出を省略し、数14に従って、 $E[i]$ の局所ピークを判定し、これをアタック時刻としても良い。この場合には、時間変化量 $\Delta E[i]$ を求める演算が不要になることから更に処理時間を短縮することができる。

【0044】

【数14】

$$E[i] > E[i-1] \cap E[i] > E[i+1] \cap E[i] > \beta$$

【0045】ただし、 $\beta$ は所定の定数であり、閾値を示す。以上説明した予測誤差算出手段3及び判定手段4は、CPUによるソフトウェア処理でも実現可能であり、従来方式と比較して計算量が大幅に少ないので、コストをあげることなく処理時間の大幅な短縮が可能であり、しかも検出精度も高く維持することができる。尚、これらの処理手順は一例であり、その主旨を逸脱しない範囲において種々の変更を加えることができる。

まず、上記数12で $i=1$ とし(S21)、 $\Delta E[i]$ を算出する(S22)。そして、この操作を $i$ がFNになるまで(S23)、 $i$ を1つずつインクリメントしながら行い(S24)、全ての $\Delta E[i]$ を求める。

【0040】次に、先に求めた上記 $\Delta E[i]$ の局所ピークの判定を行なう。図5に示すように、 $\Delta E[i]$ は正または負の値をとる。 $\Delta E[i]$ が正の大きな値をとる場所は、フレームの予測残差が急激に増加している場所であり、アタック時刻とみなせる。この判定には、数13を使い、この条件を満たすフレームの時刻をアタック時刻とする。

【0041】

【数13】

【0046】

【発明の効果】以上説明したように、本発明のアタック時刻検出装置によれば、次のように優れた作用効果を発揮することができる。従来用いていた対数パワースペクトルの代わりに線形予測法による残差を使うことにより、計算量を大幅に削減できる。従って、特別なハードウェアを付加することなく汎用CPUによるソフトウェア処理でも実用的な処理時間が得られ、低コストで検出精度の良いアタック時刻検出装置を得ることができる。

【図面の簡単な説明】

【図1】自動採譜装置等に用いられる本発明に係るアタック時刻検出装置を示すブロック構成図である。

【図2】図1の装置で作成されるフレームを説明する説明図である。

【図3】予測残差算出手段の動作を示すフローである。

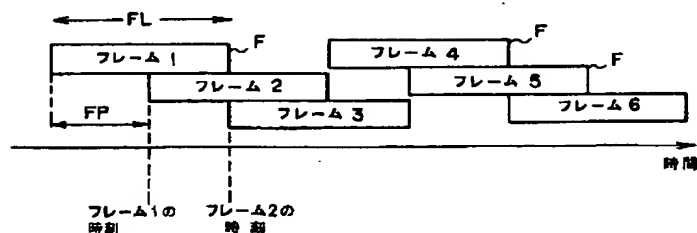
【図4】判定手段の動作を示すフローである。

【図5】アタック時刻の判定を説明するための説明図である。

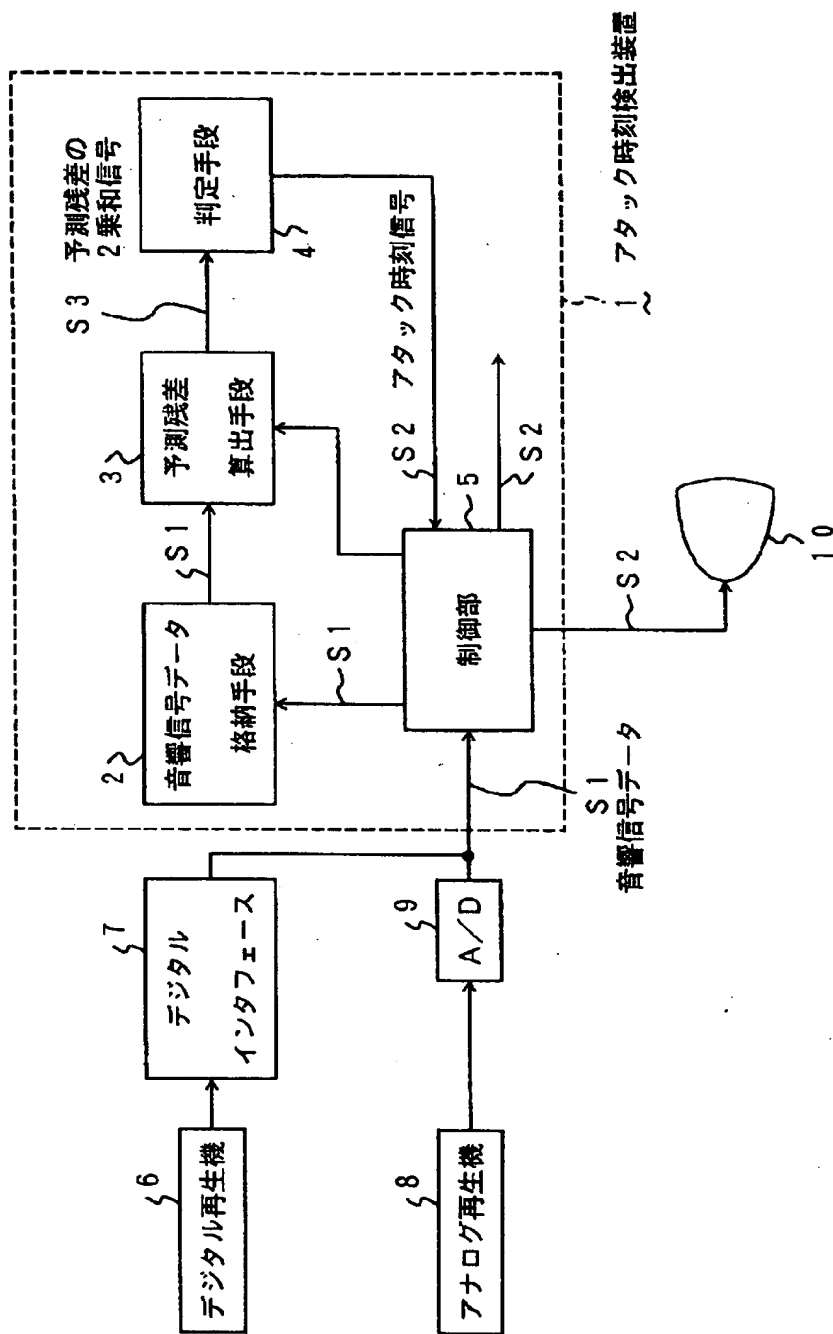
【符号の説明】

1…アタック時刻検出装置、2…音響信号データ格納手段、3…予測残差算出手段、4…判定手段、5…制御部、6…デジタル再生機、8…アナログ再生機、F…フレーム、S1…音響信号データ、S2…アタック時刻信号、S3…予測残差の2乗和信号。

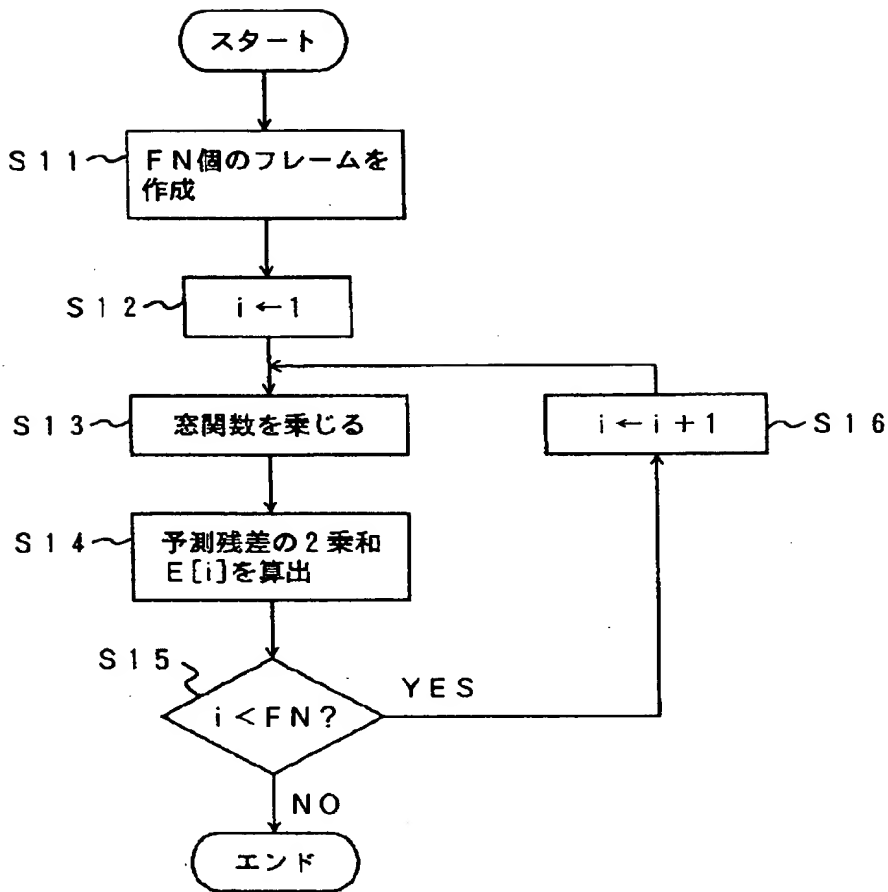
【図2】



【図1】



【図3】



【図4】

【図5】

